DOCUMENT RESUME

ED 209 329 TM 810 871

AUTHOR	Breland, Hunter M.; Griswold, Philip A.
TITLE	Group Comparisons for Basic Skills Measures.
INSTITUTION	College Entrance Examination Board, New York, N.Y.; Educational Testing Service, Princeton, N.J.
REPORT NO	CB-R-81-6; ETS-RR-81-21
PUB DATE	81
NOTE	35p.
EDRS PRICE	MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS	*Achievement Tests; Asian Americans; *Basic Skills; Blacks; *Comparative Analysis; Correlation; Higher Education; Hispanic Americans; Minority Groups; *Predictive Measurement; *Racial Differences; Regression (Statistics); *Sex Differences; Whites
IDENTIFIERS	English Placement Test; Scholastic Aptitude Test; Test of Standard Written English

ABSTRACT

Correlation, regression, and score interval analyses were conducted for academic tests including the Scholastic Aptitude Test (SAT), Test of Standard Written English (TSWE), and four subtests of the English Placement Test (EPT) for six different groups. The groups were men, women, Asians, blacks, Hispanics, and whites. The use of grade point average or individual course grades was avoided because these are subjective, often biased, measures. The EPT-Essay score based on a forty-five minute writing sample was considered a good standard for comparing other basic skills measures available. The measures were evaluated as predictors of essay writing and overall performance. The correlational comparisons showed few differences across groups, except that correlations tended to be lower for the white sample because of variance restrictions. The regression comparisons agreed with previous studies showing blacks and Hispanics to be generally overpredicted. False negatives, those people who score low on a predictor but who excel on a criterion, occurred least in the black and Hispanic groups. (Author/DWH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED209329

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

College Board Report

No. 81-6

Group Comparisons for Basic Skills Measures

Hunter M. Breland
Philip A. Griswold

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

P. K. Hendel

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Group Comparisons for Basic Skills Measures

Hunter M. Breland
Philip A. Griswold
Educational Testing Service

College Board Report No. 81-6
ETS RR No. 81-21

College Entrance Examination Board, New York, 1981

Hunter M. Breland and Philip A. Griswold are staff members of Educational Testing Service, Princeton, New Jersey.

Researchers are encouraged to express freely their professional judgment. Therefore, points of view or opinions stated in College Board Reports do not necessarily represent official College Board position or policy.

The College Board is a nonprofit membership organization that provides tests and other educational services for students, schools, and colleges. The membership is composed of more than 2,500 colleges, schools, school systems, and education associations. Representatives of the members serve on the Board of Trustees and advisory councils and committees that consider the programs of the College Board and participate in the determination of its policies and activities.

Additional copies of this report may be obtained from College Board Publication Orders, Box 2815, Princeton, New Jersey 08541. The price is $4.

Copyright © 1981 by College Entrance Examination Board.
All rights reserved.
Printed in the United States of America.

CONTENTS

Abstract iv

Introduction 1

Instruments 2

Procedures 3

Correlational Analyses 4

Regression Analyses 8

Analyses by Score Levels 8

Summary 17

References 18

Appendixes

A. Intercorrelation Matrices 20

B. Regression Analyses: Procedures and Tables 27

TABLES

1. Correlations with EPT-Essay Score 5

2. Correlations between the EPT-Reading Score and ATP Scores 5

3. Correlations between the EPT-Sentence Construction Score and ATP Scores 6

4. Correlations between the EPT-Logic and Organization Score and ATP Scores 6

5. Correlations between EPT-Total Score and ATP Scores 7

6. Predicting the EPT-Essay Score and EPT-Total Score from Multiple EPT and ATP Scores 7

7. Covariance Analyses 9

8. Essay Writing Performance by Group and TSWE Score Level 10

9. Essay Writing Performance by Group and SAT-V Score Level 12

10. Essay Writing Performance by Group and EPT-Reading Score 13

11. Essay Writing Performance by Group and EPT-Sentence Construction Score 14

12. Essay Writing Performance by Group and EPT-Organization and Logic Score 15

13. EPT-Total Score Performance by Group and TSWE Score Level 16

ABSTRACT

Correlation, regression, and score interval analyses were conducted for six academic measures as predictors of essay writing and overall performance. Comparisons for all analyses were made for men, women, Asians, blacks, Hispanics, and whites. The correlational comparisons showed few differences across groups, except that correlations tended to be lower for the white sample because of variance restrictions. The regression comparisons agreed with previous studies showing blacks and Hispanics to be generally overpredicted. On essay-writing performance, men were also overpredicted by conventional basic skills measures. In contrast, women tended to write better essays than would have been predicted by conventional basic skills measures. False negatives, those people who score low on a predictor but who excel on a criterion, occurred least in the black and Hispanic groups.

INTRODUCTION

Because of contemporary concerns about fairness to sex and ethnic groups, it has become customary to conduct analyses of academic tests in use or proposed for use to determine if any unfairness may result from such use. A previous study (Breland, 1977b) reported a number of such analyses for the Test of Standard Written English (TSWE), but this report was limited by the available data. Because of the relatively small number of minority students in the sample, it was necessary to combine all minorities into one group for analysis. The present report does not have that limitation.

The most common approach to the examination of fairness is to compare correlational and regression analyses for the various groups of interest. However, these kinds of analyses do not reveal certain kinds of potential unfairness. For example, they do not examine specific groups of people who may score low on a predictor but who perform satisfactorily if allowed to enter an educational program. Such people have been termed 'false negatives." Conversely, there exist groups of students who score high on a predictor but who do not perform satisfactorily later. These have been called "false positives." When false positives and false negatives are summed, they represent "misses" — people for whom an erroneous prediction was made. Correlational and regression analyses may not reveal the actual consequences represented in false negatives and false positives. For that reason it is important to examine the bivariate distributions of predictor scores and actual performance measures to understand better the potential for unfairness in the use of tests.

A second issue in fairness investigations is whether the outcome being predicted is reliable and unbiased. This concern is commonly referred to as the "criterion problem." The criterion usually used in prediction studies of fairness is the grade point average (GPA). Another useful criterion is performance on some task quite different from the predictor. For example, essays written in English composition courses can be used as a criterion for comparing performances for different groups on multiple-choice tests intended for the prediction of college performance. This kind of criterion has the advantage that it is usually more objective than GPA or specific college course grades — especially when essay scores are assigned by readers remote from the instructional setting.

Previous studies of some of the same basic skills measures investigated in the present study have been reported by a number of writers. Breland (1977a) reported an investigation of the Test of Standard Written English (TSWE) and its use in college English placement for four institutions. The institutions provided data on students' performance during college, and Educational Testing Service matched these data with other data obtained at the time students applied for college. Student performances on essay tests of writing ability were shown to have a strong relationship to student performance on multiple-choice tests of writing ability as represented by the TSWE. Second, the analyses indicated that a brief multiple-choice test of writing ability predicted actual writing performance during the freshman year of college as well as or better than a brief essay test given at the beginning of the freshman year.

As noted above, Breland (1977b) compared men, women, minorities, and nonminorities with respect to performance on the TSWE and subsequent performance in English composition courses as well as performance on brief impromptu essays. No important group differences in traditional correlational analyses for either grade or essay prediction were observed. Analyses of correct and incorrect placement decisions (hits and misses) at specific TSWE cutoff scores revealed no noteworthy group differences whether outcomes were based on English course grades or on freshman writing performance. From all groups, the TSWE appeared to predict freshman-year writing performance as well as or better than precourse writing samples, high school English grades, or high school rank in class. Because of the limited number of cases

available for analysis, however, it was not possible to conduct analyses within minority subgroups.

Bianchini (1977) presented some basic data for the August 1977 administration of the English Placement Test (EPT). These basic data included means, standard deviations, KR-20 reliabilities, inter-reader correlations for the essay, and speed and power analyses. Rankin (1978) visited the 19 campuses of the California State Universities and Colleges (CSUC) using the EPT and reported on its use. This study was conducted to determine how the EPT was being used at CSUC campuses. Several recommendations were made for improvements in instruction and placement. Dunbar, Minnick, and Oleson (1978) collected data for the Hayward campus for the purpose of examining the diagnostic capabilities of the EPT, especially for students with minimum writing proficiency. They conducted correlational analyses, multiple regression analyses, and discriminant analyses. These analyses indicated that the EPT essay score contributed significantly to the prediction of course performance but that the contribution was not large. Some questions were raised concerning the cost-effectiveness of the EPT as a result of high observed correlations between EPT and other measures available.

Michael and Shaffer (1978) collected data at the Northridge campus of CSUC and conducted extensive analyses. The purpose of these analyses was to examine comparative validities of the EPT and other measures available for placement as well as comparative validities for different groups. As did Dunbar et al., Michael and Shaffer suggested that there were substantial intercorrelations among tests studied. Correlations were also reported between test scores and fall-semester GPA and grades in a freshman English course. The EPT-Total score correlated .30 with GPA and .47 with English grades. These correlations were slightly higher, as would be expected, than those for the shorter SAT-V test: .27 with GPA and .41 with English grades. One would expect greater reliability, and consequently higher correlations, using the longer test. Mixed results were obtained from analyses of small groups of black and Hispanic students, as some results suggested good predictability of course performance and others did not.

In another study Bailey (1978) conducted a factor analysis in which SAT subscores, the TSWE, and EPT subscores were entered. The primary objective in this analysis was to determine if both sets of tests (the EPT and the SAT-TSWE set) would load on a common factor or whether unique properties might be possessed by either. A strong common factor was identified accounting for 80 percent of the variance from all these tests. It was suggested that all the tests have many common properties and that the EPT and the SAT (including TSWE) may be measuring similar abilities.

Instruments

This report presents comparative analyses of seven academic tests: the Scholastic Aptitude Test (with its verbal and mathematical sections considered separately), the Test of Standard Written English (TSWE), and four subtests of the English Placement Test (EPT). The four subtests are EPT-Reading, EPT-Sentence Construction, EPT-Logic and Organization, and EPT-Essay. Each of the tests is described briefly in the following paragraphs.

SAT-Verbal. This section of the SAT is made up of two verbal sections, each one requiring 30 minutes. The questions measure the ability to understand what has been read, as well as the extent of vocabulary development. SAT-Verbal scores are reported on a scale with a range of 200-800. The national mean for 1980 college-bound seniors was 424, with a standard deviation of 110.

SAT-Mathematical. This has two mathematical sections, each one requiring 30 minutes. The questions measure problem-solving abilities closely related to college work.

SAT-Mathematical scores are also reported on a 200-800 scale. The national mean for 1980 college-bound seniors was 466, with a standard deviation of 117.

TSWE. The Test of Standard Written English is a 30-minute, multiple-choice test administered with the SAT. The questions evaluate the ability to recognize standard written English, the language of most college textbooks, and the English that college students are usually expected to use in papers they write for courses. TSWE scores are reported on a scale of 20 to 60+, with 60+ representing all scores of 60 and above. The national mean for 1980 college-bound seniors was 42.4, with a standard deviation of 11.0

EPT-Reading. A 50-item, 40-minute[1] test designed to measure several reading skills such as identifying the main idea, interpreting directly or indirectly stated ideas, inferring the meaning of words within the context of a reading passage, and recognizing levels of meaning through figurative phrases. Scores have a range from 120 to 180, with a mean of 150 and a standard deviation of 10.

EPT-Sentence Construction. This is a 50-item, 40-minute test designed to measure how well students can recognize arrangements of sentence elements that express the meaning clearly and correctly. The test has the same scale as EPT-Reading.

EPT-Logic and Organization. This is a 50-item, 40-minute test designed to measure the ability to see relationships between words, sentences, objects, and ideas. The test has the same scale as EPT-Reading and EPT-Sentence Construction.

EPT-Essay. A 45-minute essay is written in response to a special topic presented. The essay is scored on a six-point scale by two readers, independently, and these two scores are added to yield a score in the range from 2 to 12. In cases of extreme disagreement between readers, a third reader is used to resolve the difference.

Procedures

This report presents comparative analyses of seven academic tests for six different groups: men, women, Asians, blacks, Hispanics, and whites.[2] A principal objective in the analyses was to avoid the use of GPA or of individual course grades as criteria, since these are subjective measures and are often biased. Because of the objective nature of writing assessment through multiple judgments by readers of essay samples, the EPT-Essay score based on a 45-minute writing sample was viewed as a good standard for comparing other basic skills measures available. Additionally, the EPT-Total score (which includes the essay) was considered a good standard for some analyses because of the large amount of information represented in it.

Correlational analyses were conducted to show basic relationships among measures for the total sample available and for subgroups as well. Regression analyses were conducted to explore further the nature of the relationships represented in summary form in the correlations. The regression analyses were essentially intended to determine whether within-group regression lines differed significantly for different groups.

Because correlational regression analyses as commonly conducted make assumptions of linearity in relationships, analyses were also made at different score levels as a check on the other analyses. The analyses within score levels addressed the question: What is the expectation of performance on a criterion measure given performance within a certain range on the predictor measure?

---

1. Recent versions of EPT tests require slightly less testing time.
2. The analyses are based on data collected by the California State Universities and Colleges (CSUC) for students entering in 1977 and matched with College Board data provided by Educational Testing Service.

Correlational Analyses

Tables 1 through 4 show correlations between the EPT components — the EPT-Essay, the EPT-Reading score, the EPT-Sentence Construction score, the EPT-Logic and Organization score — and the other available scores. Table 1 would suggest that the EPT-Essay score is about equally well predicted by the other EPT components, the TSWE, and the SAT-V. Only the SAT-M correlations with the EPT-Essay score are noticeably lower, as would be expected. This trend is consistent for all groups, except whites. Interestingly, the white sample correlations are lowest of all. The lower white correlations are at least in part attributable to the lower white standard deviation (1.71) on the EPT-Essay score. This is somewhat lower than the standard deviations for Asians (1.93), blacks (1.86), and Hispanics (1.94) on the EPT-Essay. Similar differences in variances occurred for most of the other measures studied, and these would further reduce correlations attainable for whites.[3]

All the correlations are attenuated somewhat, however, by the generally low reliability of writing sample assessment as represented in the EPT-Essay. As noted in Table 1, the reliability of a single essay scored independently by two readers was estimated at .38 by Coffman (1966). Breland and Gaynor (1979) used a longitudinal design to obtain a test-retest reliability estimate of .52 for a similar direct (writing sample) assessment. But Werts, Breland, Grandy, and Rock (1980) conducted analyses suggesting that the test-retest estimate of .52 may be an overestimate because of the potential of correlated errors of measurement resulting from the writing sample scoring. That is, readers may be systematically influenced by extraneous factors (possibly handwriting quality, neatness, and similar influences). These analyses suggest that the reliability of essay assessment with one sample and two readers is probably in the range from about .35 to about .45.

Table 2 shows that the correlational relationships increase somewhat when the multiple-choice EPT-Reading score is related to College Board scores (TSWE, SAT-V, SAT-M). SAT-V correlated best with EPT-Reading for all groups. As before, the lowest SAT-V correlations were obtained for the white sample.

In Table 3, the EPT-Sentence Construction score is correlated with the College Board scores. As with the EPT-Reading score in Table 2, these correlations are generally higher than with the EPT-Essay score. The difference between the EPT-Reading and EPT-Sentence Construction scores is that, while the former tended to correlate best with SAT-V, the latter tended to correlate best with TSWE. That is, of course, what one would expect.

Table 4 shows the correlational relationships between the EPT-Logic and Organization test and the College Board scores. Here the SAT-V appears to be the best correlate, but the white sample, as before, yielded the lowest correlations.

Table 5 relates the composite of the EPT components (the EPT-Total score) and the College Board scores. These correlations exhibit the same pattern as obtained for the EPT components but are generally higher because of the greater reliability of the EPT-Total score.

In Table 6, EPT and College Board scores are first separately combined to generate a comparative prediction of the essay score for all groups. Second, the EPT-Total score is predicted from the multiple of the College Board scores. These weighted composites appear to predict the EPT-Essay criterion equally well, with only minor variations for some groups. The white sample essay performance is by far the most difficult to predict, and this finding is consistent with all the other correlational analyses. In the analyses of Table 6, it should be noted that SAT-M

3. See Appendix A for further details. Note also that in a number of the tables in this report figures for ethnic groups do not add up to the total because identification is not available for all cases.

TABLE 1. Correlations with the EPT-Essay Score

Group	N	Correlations[a] between the EPT-Essay Score and: EPT-Reading Score	EPT-Sentence Construction Score	EPT-Logic and Organization Score	TSWE Score	SAT-V Score	SAT-M Score
Total	10,674	.46	.49	.46	.49	.45	.23
Men	4,766	.43	.48	.45	.48	.44	.29
Women	5,908	.49	.51	.48	.50	.49	.32
Asian	606	.57	.59	.54	.55	.52	.16
Black	583	.53	.56	.56	.56	.53	.30
Hispanic	445	.50	.55	.49	.53	.50	.28
White	5,236	.30	.35	.29	.37	.32	.12

a. Interpretation of these correlations should be made with an awareness that they are somewhat attenuated due to the low reliability of the essay criterion measure. Coffman (1966) has estimated that the score reliability of a single essay scored independently by two readers (the case for these data) is about .38. Corrections for attenuation for reliability in the criterion would increase a correlation of .50, for example, to .81.

TABLE 2. Correlations between the EPT-Reading Score and ATP[a] Scores

Group	N	ATP[a] Scores TSWE	SAT-V	SAT-M
Total	10,674	.67	.74	.46
Men	4,766	.64	.72	.46
Women	5,908	.69	.75	.50
Asian	606	.73	.80	.29
Black	583	.66	.76	.51
Hispanic	445	.66	.73	.46
White	5,236	.55	.65	.36

a. The Admissions Testing Program of the College Board, represented here by SAT-V, SAT-M, and TSWE.

TABLE 3. Correlations between the EPT-Sentence Construction Score and ATP[a] Scores

Group	N	ATP[a] Scores		
		TSWE	SAT-V	SAT-M
Total	10,674	.74	.70	.49
Men	4,766	.72	.68	.51
Women	5,908	.76	.72	.54
Asian	606	.77	.78	.39
Black	583	.75	.71	.54
Hispanic	445	.71	.72	.53
White	5,236	.66	.60	.39

a. The Admissions Testing Program of the College Board, represented here by SAT-V, SAT-M, and TSWE.

TABLE 4. Correlations between the EPT-Logic and Organization Score and ATP[a] Scores

Group	N	ATP[a] Scores		
		TSWE	SAT-V	SAT-M
Total	10,674	.67	.71	.52
Men	4,766	.64	.70	.53
Women	5,908	.69	.73	.54
Asian	606	.73	.77	.38
Black	583	.68	.77	.60
Hispanic	445	.64	.71	.51
White	5,236	.55	.61	.42

a. The Admissions Testing Program of the College Board, represented here by SAT-V, SAT-M, and TSWE.

TABLE 5. Correlations between EPT-Total Score and ATP[a] Scores

Group	N	ATP[a] Scores		
		TSWE	SAT-V	SAT-M
Total	10,674	.76	.77	.51
Men	4,766	.74	.76	.54
Women	5,908	.77	.79	.56
Asian	606	.79	.82	.37
Black	583	.77	.80	.57
Hispanic	445	.74	.78	.52
White	5,236	.68	.68	.40

a. The Admissions Testing Program of the College Board, represented here by SAT-V, SAT-M, and TSWE.

TABLE 6. Predicting the EPT-Essay Score and EPT-Total Score from Multiple EPT[a] and ATP[b] Scores

Group	N	Multiple Correlation, EPT-Essay Score and:		Multiple Correlation, EPT-Total Score and ATP[b]
		EPT[a]	ATP[b]	
Total	10,674	.51	.51	.83
Men	4,776	.50	.50	.82
Women	5,408	.53	.53	.84
Asian	606	.61	.58	.86
Black	583	.60	.60	.86
Hispanic	445	.57	.56	.84
White	5,236	.37	.39	.75

a. Reading, Sentence Construction, and Logic and Organization.
b. TSWE, SAT-V, and SAT-M.

contributed very little to the multiple correlation coefficient -- usually only about one additional hundredth. The final column of Table 6 shows the substantial relationship between the EPT-Total score and the composite of the ATP components.

Regression Analyses

The above correlational analyses demonstrate the degree of relationship among the several tests, under the assumption that relationships are linear. An examination of the nature of these linear relationships, however, requires analyses of the regression systems related to the correlations.

Comparisons of regression systems for different groups were conducted through analyses of covariance. Under assumptions of linearity, regression lines are parallel (i.e., the slopes are equal) when no interaction is inferred between groups and predictors. That is, a unit increase in the predictor produces a proportionate increase in the criterion regardless of group membership. Second, if the lines are coincident (i.e., the slopes and intercepts are equal) the equation for one group is essentially the same as the equation for the other groups. The statistical procedures and other details of the covariance analyses are given in Appendix B.

The results of the covariance analyses are given in Table 7. In every ethnic comparison with whites on either TSWE or SAT-V, statistically significant interactions were found ($p < .01$). However, the small changes in the multiple R's shown in Table 7, as group and interaction influences were added, suggest that the interaction effects were small. When the nonessay components of the EPT were used to predict scores on the EPT-Essay, no significant interactions were found for blacks and whites, but tests of coincidence of the two regression lines showed significant differences ($p < .01$). This suggests that the white regression overpredicts for blacks throughout the range of EPT scores. A similar pattern is evident for prediction of Hispanic scores. Some interaction is present, but the white regression tends to overpredict Hispanic performance. The pattern for prediction of Asian scores is the least distinct, although significant interactions do exist. Again, the multiple R's suggest little influence of group or interaction factors.

The results for the regression comparison of men and women are quite clear. No interaction exists for TSWE, SAT-V, or the EPT component predictors. Since the slopes are parallel and all the tests of coincidence are significant ($p < .01$), the regression suggests that women are higher at all score levels than men. Thus, the essay writing performance of women would tend to be underpredicted (and that of men overpredicted) by a linear equation based on data for both sexes combined.

Analyses by Score Levels

While the regression comparisons just described were not indicative of important group differences, the statistically significant interactions and intercept differences are intriguing and worthy of further exploration. One approach toward exploring the nature of any group differences is to conduct analyses within score levels. Such a procedure allows for the detection of nonlinearities in the data and also provides a sense of the practical importance of group differences.

Table 8 presents group comparisons of EPT-Essay performance for four different score intervals of the TSWE. The percentages shown at the bottom of Table 8 are the observed proportions of group members within a given score range who wrote better-than-average essays. Tests of significance were based on the percentages for the total frequencies at the left. Thus, for example, it would be expected that 70.3 percent of those in the sample scoring 50 or above on the TSWE would write better-than-average essays. The tests of significance show that women scoring 50

TABLE 7. Covariance Analyses

Groups Compared	Regressions Compared	N	R Using Basic Predictor	R Adding Group	R Adding Interaction	F for Interaction (Test of Slope)	F for Coincidence[a] (Test of Intercept)
Blacks vs. whites	Essay on TSWE	5,819	.44	.45	.46	25.4*	-
Blacks vs. whites	Essay on SAT-V	5,819	.40	.41	.42	42.8*	-
Blacks vs. whites	Essay on EPT-R	5,819	.40	.40	.40	1.8	15.0*
Blacks vs. whites	Essay on EPT-SC	5,819	.44	.44	.44	1.3	11.9*
Blacks vs. whites	Essay on EPT-L&O	5,819	.40	.40	.40	1.9	8.4*
Hispanics vs. whites	Essay on TSWE	5,681	.41	.41	.42	19.6*	-
Hispanics vs. whites	Essay on SAT-V	5,681	.36	.37	.38	28.6*	-
Hispanics vs. whites	Essay on EPT-R	5,681	.35	.35	.36	5.6	18.5*
Hispanics vs. whites	Essay on EPT-SC	5,681	.40	.40	.41	5.1	13.5*
Hispanics vs. whites	Essay on EPT-L&O	5,681	.34	.35	.35	3.8	15.3*
Asians vs. whites	Essay on TSWE	5,842	.41	.42	.42	13.0*	-
Asians vs. whites	Essay on SAT-V	5,842	.36	.37	.38	25.7*	-
Asians vs. whites	Essay on EPT-R	5,842	.37	.37	.37	1[illegible] 5*	-
Asians vs. whites	Essay on EPT-SC	5,842	.41	.41	.41	4.9	7.5*
Asians vs. whites	Essay on EPT-L&O	5,842	.35	.36	.36	7.2*	-
Blacks vs. Hispanics	Essay on TSWE	1,028	.56	.56	.56	<1	<1
Blacks vs. Hispanics	Essay on SAT-V	1,028	.53	.53	.53	<1	1.0
Blacks vs. Hispanics	Essay on EPT-R	1,028	.52	.52	.52	<1	1.2
Blacks vs. Hispanics	Essay on EPT-SC	1,028	.56	.56	.56	<1	<1
Blacks vs. Hispanics	Essay on EPT-L&O	1,028	.53	.53	.53	<1	2.3
Men vs. women	Essay on TSWE	10,674	.49	.51	.51	<1	171.1*
Men vs. women	Essay on SAT-V	10,674	.45	.49	.49	<1	279.3*
Men vs. women	Essay on EPT-R	10,674	.46	.49	.49	<1	224.4*
Men vs. women	Essay on EPT-SC	10,674	.49	.52	.52	<1	193.0*
Men vs. women	Essay on EPT-L&O	10,674	.46	.49	.49	<1	242.9*

a. Where dashes are indicated, the standard practice was followed of not performing tests of coincidence when the interactions are significant.

*$p<.01$ (Although significance occurs among the groups, note that little increase in R was found. This may be due to slight nonlinearity of the data among the first three groups and, in the last group, to the very large N.)

TABLE 8. Essay Writing Performance by Group and TSWE Score Level

TSWE Score	Total	Men	Women	Asian	Black	Hispanic	White
	Frequencies Scoring in Four TSWE Ranges						
50+	2,855	1,171	1,684	87	33	35	1,738
40-49	3,914	1,724	2,190	206	113	146	2,131
30-39	2,659	1,300	1,359	176	196	159	1,115
Below 30	1,246	571	675	137	241	105	252
	Frequencies Writing Above-Average Essays						
50+	2,006	732	1,284	51	23	26	1,229
40-49	2,187	776	1,311	101	62	80	1,136
30-39	850	341	509	60	40	32	408
Below 30	167	58	109	15	22	14	58
	Percentages Writing Above-Average Essays						
50+	70.3	62.5*	75.6*	58.6*	69.7	74.3	70.7
40-49	53.3	45.0*	59.9*	49.0	54.9	54.8	53.3
30-39	32.0	26.2*	37.4*	34.1	20.4*	20.1*	36.6*
Below 30	13.4	10.2*	16.1*	10.9	9.1*	13.3	23.0*

*Statistically significant ($p < .05$) deviation from expected percentage.

or above on TSWE wrote significantly more above-average essays (75.6 percent) than expected and that men and Asians who scored 50 or above on the TSWE wrote significantly fewer above-average essays than expected. The differences between men and women, moreover, occurred at all four TSWE score levels. Blacks and Hispanics scoring below 40 on TSWE tended to write fewer above-average essays than expected, though the 13.3 percent figure for Hispanics in the lowest (below 30) range was not statistically significant.

Conversely, whites scoring below 40 on the TSWE tended to write more above-average essays than would be expected for people in those score ranges. The analyses represented in Table 8 are useful as explanations of the interaction and intercept differences reported in Table 7. The differences between men and women are consistent and are highlighted by Table 8, whereas the multiple R's of Table 7 tended to mask them.

Table 9 shows similar analyses for the SAT-V as a predictor. The pattern is quite similar to that observed for the TSWE in Table 8. Women consistently write better essays than men in the same SAT-V score intervals, and whites who score below 400 on the SAT-V write better essays than blacks and Hispanics with SAT-V scores in the same intervals.

Tables 10, 11, and 12 present comparable analyses for EPT-Reading, EPT-Sentence Construction, and EPT-Logic and Organization as predictors of essay writing performance. Like the TSWE and SAT-V, the EPT components tend to underpredict essay performance for women and to overpredict for men. And blacks and Hispanics scoring in the highest EPT intervals tend to write fewer above-average essays than would be expected from their multiple-choice test performance.

Tables 8 through 12 may also be considered as analyses of false negatives and false positives. Those scoring low on a predictor but writing above-average essays are false negatives, and those scoring high on a predictor but writing below-average essays are false positives. In Table 8, for example, the highest proportions of false negatives predicted by TSWE are for women and whites, and the lowest proportions for men, blacks, and Hispanics. In Table 9 the same pattern is shown with the SAT-V as a predictor. Tables 10 through 12 also show this pattern. One should note, however, that what is "low" for the EPT components is different from what is low for SAT-V and TSWE. The average score for the EPT components is about 150, so that most of the figures in Tables 10, 11, and 12 are for relatively low EPT scores.

Another way of viewing the issue of false negatives is to consider it in connection with the EPT-Total score. Even though the correlational comparisons presented previously show substantial relationships between the EPT-Total score and the other measures, such analyses do not examine performance by group within specific score ranges. Since the previous analyses indicated few differences among the various tests, it is enough to limit this last comparison to one of the tests. Table 13 presents a comparison of TSWE and EPT-Total scores for all the groups. This table illustrates the strong relationship between these two tests, which was demonstrated before by the high (.76) correlation reported in Table 5. For those students scoring 50 or above on the TSWE, 96.0 percent were above average (150) on the EPT-Total score. There are some variations across groups, with only 87.9 percent of the black sample with high TSWE scores scoring above average on the EPT-Total, but these differences are not statistically significant. At the other extreme of the TSWE distribution (below 30), only 6.9 percent of those students entering CSUC in 1977 obtained above-average EPT-Total scores. Blacks scoring below 30 on the TSWE seem least likely to perform well on the EPT, since only 2.5 percent of blacks scoring below 30 on the TSWE had above-average EPT-Total scores. These comparisons of EPT-Total scores and TSWE scores lead to essentially the same conclusion as the previous analyses. The smallest proportions of false negatives tend to occur for the black and Hispanic groups. The one noticeable difference between the pattern of Table 13 and that of other comparisons is that there were no statistically significant differences between men and women.

TABLE 9. Essay Writing Performance by Group and SAT-V Score Level

SAT-V Score	Total	Men	Women	Asian	Black	Hispanic	White
	Frequencies Scoring in Four SAT-V Ranges						
500+	2,378	1,093	1,285	85	18	38	1,442
400-490	4,246	2,000	2,246	208	93	126	2,332
300-390	3,076	1,330	1,746	196	238	191	1,315
Below 300	974	343	631	117	234	90	147
	Frequencies Writing Above-Average Essays						
500+	1,640	651	989	53	10	30	1,004
400-490	2,257	876	1,381	103	50	62	1,256
300-390	1,105	353	752	66	70	56	543
Below 300	108	27	81	11	17	7	34
	Percentages Writing Above-Average Essays						
500+	69.0	59.6*	77.0*	67.3	55.6	78.9	69.6
400-490	53.2	43.8*	61.5*	49.5	53.8	49.2	53.8
300-390	35.9	26.5*	43.1*	33.7	29.4*	29.3*	41.3*
Below 300	11.1	7.9*	12.8	9.4	7.3*	7.8	23.1*

*Statistically significant ($p < .05$) deviation from expected percentage.

TABLE 10. Essay Writing Performance by Group and EPT-Reading Score

EPT-Reading Score	Total	Men	Women	Asians	Blacks	Hispanics	Whites
	Frequencies Scoring in Four EPT-Reading Score Ranges						
150+	6,522	2,934	3,588	272	132	165	3,773
140-149	2,978	1,352	1,626	192	203	191	1,262
130-139	716	294	422	75	120	52	168
Below 130	458	186	272	67	128	37	33
	Frequencies Writing Above-Average Essays						
150+	3,839	1,449	2,390	164	64	92	2,267
140-149	1,116	407	709	60	60	54	521
130-139	133	47	86	17	13	7	43
Below 130	22	4	18	2	5	2	6
	Percentages Writing Above-Average Essays						
150+	58.9	49.4*	66.6*	60.3	48.5*	55.8	60.1
140-149	37.5	30.1*	43.6*	31.2	29.6*	28.3	41.3
130-139	18.6	16.0	20.4	22.7	10.8*	13.5	25.6*
Below 130	4.8	2.2	6.6	3.0	3.9	5.4	18.2*

*Statistically significant ($p < .05$) deviation from expected percentage.

TABLE 11. Essay Writing Performance by Group and EPT-Sentence Construction Score

EPT-Sentence Construction Score	Total	Men	Women	Asians	Blacks	Hispanics	Whites
	Frequencies Scoring in Four EPT-Reading Score Ranges						
150+	7,050	3,092	3,958	307	166	200	4,006
140-149	2,279	1,083	1,196	154	163	131	979
130-139	900	410	490	77	142	78	203
Below 130	445	181	264	68	112	36	48
	Frequencies Writing Above-Average Essays						
150+	4,171	1,571	2,600	170	84	106	2,527
140-149	758	279	479	47	44	38	350
130-139	155	49	106	9	14	9	53
Below 130	26	8	18	7	5	2	5
	Percentages Writing Above-Average Essays						
150+	59.2	50.4*	64.2*	55.3	50.6*	53.0	63.1*
140-149	33.3	25.8*	46.9*	30.5	27.0	29.0	35.8
130-139	17.2	12.0*	22.8*	11.7	9.8*	11.5	26.1*
Below 130	5.8	4.4	6.8	10.3	4.5	5.6	10.4

*Statistically significant ($p < .05$) deviation from expected percentage.

TABLE 12. Essay Writing Performance by Group and EPT-Organization and Logic Score

EPT-Organization and Logic Score	Total	Men	Women	Asians	Blacks	Hispanics	Whites
	Frequencies Scoring in Four EPT-Reading Score Ranges						
150+	7,114	3,210	3,904	344	143	201	4,024
140-149	2,237	1,015	1,222	135	152	139	961
130-139	781	325	456	62	123	61	202
Below 130	542	216	326	65	165	44	49
	Frequencies Writing Above-Average Essays						
150+	4,105	1,554	2,551	171	73	97	2,390
140-149	838	308	530	42	53	46	396
130-139	131	36	95	11	11	3	44
Below 130	36	9	27	4	10	6	7
	Percentages Writing Above-Average Essays						
150+	57.7	48.4*	65.3*	49.7*	51.0	48.2*	59.4
140-149	37.5	30.3*	43.4*	31.1	34.9	33.1	41.2
130-139	16.8	11.1*	20.8*	17.7	8.9*	4.9*	21.8
Below 130	6.6	4.2	8.3	6.2	6.1	13.6	14.3

*Statistically significant ($p < .05$) deviation from expected percentage.

TABLE 13. EPT-Total Score Performance by Group and TSWE Score Level

TSWE Score Range	Total	Men	Women	Asian	Black	Hispanic	White
	Frequencies in Four TSWE Score Ranges						
50+	2,861	1,175	1,686	87	33	35	1,743
40-49	3,917	1,725	2,192	206	113	147	2,131
30-39	2,666	1,304	1,362	176	196	160	1,118
Below 30	1,275	587	688	148	243	107	254
	Frequencies with Above-Average EPT-Total Scores						
50+	2,748	1,119	1,629	85	29	32	1,683
40-49	3,000	1,290	1,710	143	58	102	1,695
30-39	972	449	523	60	33	38	517
Below 30	88	38	50	5	6	5	38
	Percentages with Above-Average EPT-Total Scores						
50+	96.0	95.2	96.6	97.7	87.9	91.4	96.6
40-49	76.6	74.8	78.0	69.4*	51.3*	69.4	79.5
30-39	36.4	34.4	38.4	34.1	16.8*	23.8*	46.2*
Below 30	6.9	6.5	7.3	3.4	2.5*	4.7	15.0*

*Statistically significant ($p < .05$) deviation from expected percentage.

SUMMARY

Seven tests were compared and contrasted with respect to interrelationships for six different groups. Special emphasis was placed on the use of an essay. The EPT-Essay score was used as a common criterion for making comparisons of the EPT nonessay components with the College Board's SAT-V, SAT-M, and TSWE. The EPT-Total score was also used as a criterion for some comparisons. Correlation, regression, and score interval analyses were conducted.

The correlational analyses demonstrated the close relationship among SAT-V, TSWE, and EPT scores. There were few correlational differences across groups, with the exception that the white sample consistently yielded lower correlations for all comparisons than did the other groups. Whether the relationship being examined was with the EPT-Essay score, the EPT-Reading score, the EPT-Sentence Construction score, the EPT-Logic and Organization score, or with the EPT-Total score, the correlation for the white sample was always the lowest. The lower white correlations were attributed to attenuated variances in the white sample for most scores. A second interesting observation was that of the relationship between the success of multiple components of the EPT and College Board's tests as predictors with the EPT-Essay score as a common criterion. Both sets of components predicted EPT-Essay performance equally well for all groups. Finally, the EPT-Total score was seen to be well predicted by the multiple set of College Board scores (SAT-V, SAT-M, and TSWE), even though the SAT-M contributed only minimally to this prediction.

The regression analyses indicated the nature of relationships among the various test scores. The EPT-Essay score was used as a common criterion, and all groups were compared and contrasted in regard to the nature of the predictive relationship for each of the nonessay components studied. Both the TSWE and the SAT-V were seen to overpredict minority performance in essay writing. Similarly, the EPT nonessay components tended to overpredict minority performance in essay writing. The observation of overprediction of minority performance has been a common one in previous studies of all types of tests. The regression comparisons for men and women also followed a pattern often observed in previous studies. Throughout the score ranges for TSWE, SAT-V, and all three EPT nonessay components, women were consistently underpredicted. In other words, women consistently performed better on essay writing tasks than the test scores in this analysis indicate.

The score interval analyses further illustrated the predictive relationships. The EPT-Total score appeared to be quite closely related to TSWE scores, as 96 percent of those scoring 50 or above on TSWE obtained above-average EPT-Total scores; but only 7 percent of those scoring below 30 on TSWE obtained above-average EPT-Total scores. These descriptive comparisons were similar for all groups. Distributions were also used to determine the proportions of false negatives for different groups in different TSWE score ranges, using both the EPT-Essay and the EPT-Total score as criteria. These analyses showed that blacks and Hispanics tended to have the smallest representations of false negatives, while women had the most. The results obtained were in agreement with similar, previous studies. One interesting new finding was observed, however. Although women consistently out-performed men on essay writing, they did not do so when the criterion consisted primarily of multiple-choice measures, as in the EPT-Total score.

REFERENCES

Bailey, R. L. "Principal Components Analysis of the Old and the New: The SAT vs. the CSUC English Placement Test," Measurement and Evaluation in Guidance, 11:3 (1978): 162-168.

Bianchini, J. C. The California State Universities and Colleges English Placement Test (Educational Testing Service Statistical Report SR-77-70). Princeton: Educational Testing Service, August 1977.

Breland, H. M. A Study of College English Placement and the Test of Standard Written English (ETS Project Report PR-77-1, College Board Research and Development Report RDR-76-77-4). Princeton: Educational Testing Service, January 1977a.

Breland, H. M. Group Comparisons for the Test of Standard Written English (ETS Research Bulletin RB-77-[illegible]). Princeton: Educational Testing Service, August 1977b.

Breland, H. M., and J. L. Gaynor. "A Comparison of Direct and Indirect Assessments of Writing Skill," Journal of Educational Measurement, 16:2 (Summer 1979): 119-128.

Cleary, T. A. "Test Bias: Prediction of Grades of Negro and White Students in Integrated Colleges," Journal of Educational Measurement, 5:2 (1968): 115-124.

Coffman, W. E. "On the Validity of Essay Tests of Achievement," Journal of Educational Measurement, 3:2 (Summer 1966): 151-156.

Dunbar, C. R., L. Minnick, and S. Oleson. Validation of the English Placement Test at CSUH (Student Services Report No. 2-78/79). Hayward: California State University, October 20, 1978.

Gulliksen, H., and S. S. Wilks. "Regression Tests for Several Samples," Psychometrika, 15:2 (1950): 91-114.

Humphreys, L. "Implications of Group Differences for Test Interpretation," Proceedings of 1972 Invitational Conference on Testing Problems. Princeton: Educational Testing Service, 1973, pp. 56-71.

Kleinbaum, D.G., and L. L. Kupper. Applied Regression Analysis and Other Multivariable Methods. North Scituate, Massachusetts: Duxbury Press, 1978.

Michael, W. B., and P. Shaffer. "The Comparative Validity of the California State Universities and Colleges English Placement Test (CSUC-EPT) in the Prediction of Fall Semester Grade Point Average and English Course Grades of First-Semester Entering Freshmen," Educational and Psychological Measurement, 38:4 (1978): 985-1001.

Michael, W. B., and P. Shaffer. "A Comparison of the Validity of the Test of Standard Written English (TSWE) and of the California State Universities and Colleges English Placement Test (CSUC-EPT) in the Prediction of Grades in a Basic English Composition Course and of Overall Freshman-Year Grade Point Average," Educational and Psychological Measurement, 39:1 (1979): 131-145.

Murphy, N. C. A Preliminary Validation Study, the English Placement Test, California State Universities and Colleges System, Academic Year 1977-78. San Luis Obispo: California Polytechnic State University, September 1978.

Nie, N. H., C. H. Hull, J. G. Jenkins, K. Steinbrenner, and D. H. Bent. Statistical Package for the Social Sciences, 2nd Edition. New York: McGraw-Hill Book Company, 1975.

Overall, J. E., and D. K. Spiegel. "Concerning Least Squares Analysis of Experimental Data," Psychological Bulletin, 72:5 (1969): 311-327.

Rankin, D. The Reception, Applications, and Impact of the English Placement Test in the California State Universities and Colleges. Long Beach, California: Office of the Chancellor, August 1978.

Reschly, D. J., and D. L. Sabers. "Analysis of Test Bias in Four Groups with the Regression Definition," Journal of Educational Measurement, 16:1 (1979): 1-9.

Temp, G. "Validity of the SAT for Blacks and Whites in Thirteen Integrated Institutions," Journal of Educational Measurement, 8:4 (1971): 245-251.

Werts, C. E., H. M. Breland, J. Grandy, and D. Rock. "Using Longitudinal Data to Estimate Reliability in the Presence of Correlated Measurement Errors," Educational and Psychological Measurement, 40:1 (1980): 19-29.

TABLE A-1. Intercorrelations of EPT and ATP Components for Total Sample (N=10,674)

Test	Mean	Standard Deviation	1	2	3	4	5	6	7	8
1. EPT-Essay	7.31	1.85	1.00	.46	.49	.46	.71	.45	.23	.49
2. EPT-Reading	150.17	9.14	.46	1.00	.76	.79	.89	.74	.46	.67
3. EPT-Sentence Construction	150.40	9.32	.49	.76	1.00	.76	.89	.70	.49	.74
4. EPT-Logic	150.10	9.36	.46	.79	.76	1.00	.89	.72	.52	.67
5. EPT-Total	150.18	7.84	.71	.89	.89	.89	1.00	.77	.51	.76
6. SAT-V[a]	42.42	9.50	.45	.74	.70	.72	.77	1.00	.54	.72
7. SAT-M[a]	47.40	10.64	.23	.46	.49	.52	.51	.54	1.00	.49
8. TSWE	42.44	9.95	.49	.67	.74	.67	.76	.72	.49	1.00

a. Note that the scale used for these purposes has been truncated to 20 to 80, rather than the usual 200 to 800 scale for reported scores.

TABLE A-2. Intercorrelations of EPT and ATP Components for Men (N=4,766)

Test	Mean	Standard Deviation	1	2	3	4	5	6	7	8
1. EPT-Essay	6.95	1.83	1.00	.43	.48	.45	.71	.44	.29	.48
2. EPT-Reading	150.22	8.85	.43	1.00	.74	.78	.88	.72	.46	.64
3. EPT-Sentence Construction	150.13	9.08	.48	.74	1.00	.74	.88	.68	.51	.72
4. EPT-Logic	150.33	9.10	.45	.78	.74	1.00	.88	.70	.53	.64
5. EPT-Total	149.73	7.56	.71	.88	.88	.88	1.00	.76	.54	.74
6 SAT-V[a]	42.96	9.21	.44	.72	.68	.70	.76	1.00	.53	.71
7. SAT-M[a]	50.83	10.69	.29	.46	.51	.53	.54	.53	1.00	.51
8. TSWE	41.94	9.82	.48	.64	.72	.64	.74	.71	.51	1.00

a. Note that the scale used for these purposes has been truncated to 20 to 80, rather than the usual 200 to 800 scale for reported scores.

TABLE A-3. Intercorrelations of EPT and ATP Components for Women (N=5,908)

Test	Mean	Standard Deviation	1	2	3	4	5	6	7	8
1. EPT-Essay	7.60	1.82	1.00	.49	.51	.49	.72	.49	.32	.50
2. EPT-Reading	150.12	9.37	.49	1.00	.78	.81	.90	.75	.50	.69
3. EPT-Sentence Construction	150.63	9.50	.51	.78	1.00	.78	.90	.72	.54	.76
4. EPT-Logic	149.92	9.56	.49	.81	.78	1.00	.90	.73	.54	.69
5. EPT-Total	150.54	8.03	.72	.90	.90	.90	1.00	.79	.56	.77
6. SAT-V[a]	41.98	97.01	.49	.75	.72	.73	.79	1.00	.57	.74
7. SAT-M[a]	44.63	97.56	.32	.50	.54	.54	.56	.57	1.00	.54
8. TSWE	42.85	10.05	.50	.69	.76	.69	.77	.74	.54	1.00

a. Note that the scale used for these purposes has been truncated to 20 to 80, rather than the usual 200 to 800 scale for reported scores.

TABLE A-4. Intercorrelations of EPT and ATP Components for Asian Sample (N=606)

Test	Mean	Standard Deviation	1	2	3	4	5	6	7	8
1. EPT-Essay	6.82	1.93	1.00	.57	.59	.54	.75	.52	.16	.55
2. EPT-Reading	145.82	11.14	.57	1.00	.82	.85	.92	.81	.35	.73
3. EPT-Sentence Construction	146.23	11.39	.59	.82	1.00	.83	.93	.78	.39	.77
4. EPT-Logic	147.08	11.21	.54	.85	.83	1.00	.92	.77	.38	.73
5. EPT-Total	146.68	9.61	.75	.92	.93	.92	1.00	.82	.37	.79
6. SAT-V[a]	38.90	9.89	.52	.81	.78	.77	.82	1.00	.46	.76
7. SAT-M[a]	49.27	9.81	.16	.35	.39	.38	.37	.46	1.00	.41
8. TSWE	38.16	10.47	.55	.73	.77	.73	.79	.76	.41	1.00

a. Note that the scale used for these purposes has been truncated to 20 to 80, rather than the usual 200 to 800 scale for reported scores.

TABLE A-5. Intercorrelations of EPT and ATP Components for Black Sample (N=583)

Test	Mean	Standard Deviation	1	2	3	4	5	6	7	8
1. EPT-Essay	6.20	1.86	1.00	.53	.57	.56	.74	.53	.30	.56
2. EPT-Reading	139.91	11.58	.53	1.00	.76	.80	.90	.76	.51	.66
3. EPT-Sentence Construction	140.68	11.25	.57	.76	1.00	.75	.89	.71	.54	.75
4. EPT-Logic	138.31	12.24	.56	.80	.75	1.00	.91	.77	.60	.68
5. EPT-Total	140.84	9.65	.74	.90	.89	.91	1.00	.80	.57	.77
6. SAT-V[a]	32.40	8.25	.53	.76	.71	.77	.80	1.00	.61	.69
7. SAT-M[a]	34.84	7.91	.30	.51	.54	.60	.57	.61	1.00	.55
8. TSWE	32.79	9.46	.56	.66	.75	.68	.77	.69	.55	1.00

a. Note that the scale used for these purposes has been truncated to 20 to 80, rather than the usual 200 to 800 scale for reported scores.

TABLE A-6. Intercorrelations of EPT and ATP Components for Hispanic Sample (N=445)

Test	Mean	Standard Deviation	1	2	3	4	5	6	7	8
1. EPT-Essay	6.60	1.94	1.00	.49	.55	.49	.74	.50	.28	.53
2. EPT-Reading	145.61	9.93	.49	1.00	.75	.77	.88	.73	.47	.66
3. EPT-Sentence Construction	145.53	10.30	.55	.74	1.00	.76	.90	.72	.53	.71
4. EPT-Logic	145.34	10.49	.49	.77	.76	1.00	.89	.71	.51	.64
5. EPT-Total	145.74	8.62	.74	.88	.90	.89	1.00	.76	.52	.75
6. SAT-V[a]	37.02	8.45	.50	.73	.72	.71	.78	1.00	.50	.70
7. SAT-M[a]	40.13	8.90	.28	.47	.53	.51	.52	.50	1.00	.45
8. TSWE	36.74	9.40	.53	.66	.71	.64	.75	.70	.45	1.00

a. Note that the scale used for these purposes has been truncated to 20 to 80, rather than the usual 200 to 800 scale for reported scores.

TABLE A-7. Intercorrelations of EPT and ATP Components for White Sample (N=5,236)

Test	Mean	Standard Deviation	1	2	3	4	5	6	7	8
1. EPT-Essay	7.64	1.71	1.00	.30	.35	.29	.67	.32	.12	.37
2. EPT-Reading	152.74	6.63	.30	1.00	.63	.66	.81	.65	.36	.55
3. EPT-Sentence Construction	152.94	7.06	.35	.63	1.00	.64	.83	.60	.39	.66
4. EPT-Logic	152.69	6.58	.29	.66	.64	1.00	.81	.61	.42	.55
5. EPT-Total	152.51	5.55	.67	.81	.83	.81	1.00	.69	.40	.68
6. SAT-V[a]	44.72	8.50	.32	.65	.60	.61	.69	1.00	.45	.64
7. SAT-M[a]	49.32	9.82	.12	.36	.39	.42	.40	.45	1.00	.39
8. TSWE	45.01	8.69	.37	.55	.66	.55	.68	.64	.39	1.00

a. Note that the scale used for these purposes has been truncated to 20 to 80, rather than the usual 200 to 800 scale for reported scores.

APPENDIX B. REGRESSION ANALYSES: PROCEDURES AND TABLES

Previous research comparing regression systems in group comparisons often has not specified precisely how the comparisons were made. This note is intended to rectify that situation — at least for this report.

If the regression of a criterion on a test score is the same for different groups, the test is an unbiased predictor of the criterion for those groups. In the present study, to determine the validity of predicting the EPT-Essay score from TSWE, SAT-V, and EPT component scores for various ethnic groups and for men and women, multiple regression analysis was utilized. The procedure is similar to the rationale of Gulliksen and Wilks (1950), which is based on evaluation of the quality of errors of estimate, of slopes, and of intercepts. This general methodology has been frequently used in test bias studies (e.g., Cleary, 1968; Humphreys, 1973; Reschley and Sabers, 1979; Temp, 1971).

The general linear model technique of the software provided by the Statistical Package for the Social Sciences (Nie, Hull, Jenkins, Steinbrenner, and Bent, 1975, pp. 381-383) was used because of its convenience and availability to most researchers. Since multiple regression may include dummy variables as predictors, pairs of ethnic groups to be compared were used as dichotomous, categorical variables. A least squares approach to analysis of covariance similar to the method employed by Cleary (1968) permitted testing of effects of the interaction of test score by ethnicity and of the effect of ethnicity, holding test score constant.

The single multiple regression for the two group categories (Z = 0,1), score on TSWE (X), the EPT-Essay criterion (Y'), and the interaction between race and TSWE (ZX) was developed from the rationale of Kleinbaum and Kupper (1978, Chapter 13) as follows:

$$Y' = B_0 + B_1X + B_2Z + B_3(ZX) + \text{Error}.$$

For blacks (Z = 1) the model reduces to:

$$Y_B' = (B_0 + B_2) + (B_1 + B_3)\ X + \text{Error};$$

and for whites (Z = 0):

$$Y_W' = B_0 + B_1X + \text{Error}.$$

Thus, one can test the slopes and intercepts of two groups within a single model.

Humphreys (1973) has stated that the critical comparisons to be made are tests of equal slopes and intercepts. Following his suggestion the two hypotheses tested were from the single regression equation.

1. The two regression lines are parallel; that is, H_0: $B_3 = 0$, i.e., X is proportional from group to group. When $B_3 = 0$, the slope for blacks becomes the slope for whites.

2. The two regression lines are coincident; i.e., H_0: $B_2 = B_3 = 0$, i.e., X is identical after the differences in groups are removed. When $B_2 = B_3 = 0$, the model for blacks reduces to the model for whites.

To test hypothesis 1 the following F statistic was computed:

$$F(XZ|X,Z) = \frac{SS_{\text{regression}}\ (X,Z,\ XZ) - SS_{\text{regression}}\ (X,Z)}{MS_{\text{residual}}\ (X,Z,\ XZ)}$$

To test hypothesis 2:

$$F(XZ,Z|X) = \frac{[SS_{\text{regression}}\ X,Z,\ XZ) - SS_{\text{regression}}\ (X)]/2df}{MS_{\text{residual}}\ (X,Z,\ XZ)}$$

If the slopes were significantly different, no test of coincidence of lines was conducted.

The multiple R^2 for (X, Z, XZ) and its various subsets may also be used, as it represents the proportion of total sums of squares (Overall and Spiegel, 1969).

TABLE B-1. Estimates of Simple Regression Parameters - Ethnic Comparisons

Groups	EPT-Essay on:	Intercept (B_0)	Raw Coefficient (B_1)	Standardized Coefficient	S.E. of B_1
Asian	TSWE	2.96	.1011	.547	.0063
	SAT-V	2.88	.0101	.517	.0007
	EPT-R	-7.72	.0997	.574	.0058
	EPT-SC	-7.82	.1001	.589	.0056
	EPT-L&O	-6.97	.0937	.543	.0059
Black	TSWE	2.59	.1100	.560	.0068
	SAT-V	2.30	.0120	.535	.0008
	EPT-R	-5.77	.0856	.533	.0056
	EPT-SC	-6.92	.0933	.565	.0056
	EPT-L&O	-5.47	.0844	.556	.0052
Hispanic	TSWE	2.56	.1101	.533	.0083
	SAT-V	2.32	.0115	.503	.0009
	EPT-R	-7.47	.0966	.495	.0081
	EPT-SC	-8.49	.1037	.551	.0075
	EPT-L&O	-6.67	.0913	.494	.0076
White	TSWE	4.36	.0728	.370	.0025
	SAT-V	4.74	.0065	.322	.0003
	EPT-R	-4.05	.0765	.297	.0034
	EPT-SC	-5.47	.0857	.354	.0031
	EPT-L&O	-3.87	.0754	.290	.0034

TABLE B-2. Estimates of Simple Regression Parameters - Sex Comparisons

Groups	EPT-Essay on:	Intercept (B_0)	Raw Coefficient (B_1)	Standardized Coefficient	S.E. of B_1
Females	TSWE	3.72	.0907	.500	.0020
	SAT-V	3.76	.0092	.487	.0002
	EPT-R	-6.70	.0953	.490	.0022
	EPT-SC	-7.03	.0972	.506	.0022
	EPT-L&O	-6.28	.0926	.486	.0022
Males	TSWE	3.20	.0894	.480	.0024
	SAT-V	3.16	.0088	.444	.0003
	EPT-R	-6.50	.0895	.434	.0027
	EPT-SC	-7.69	.0975	.485	.0026
	EPT-L&O	-6.53	.0897	.446	.0026